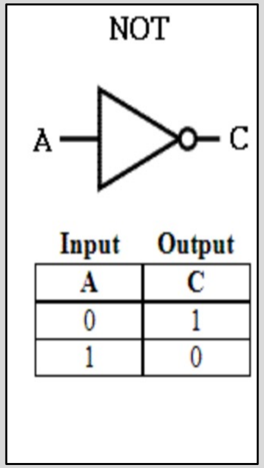
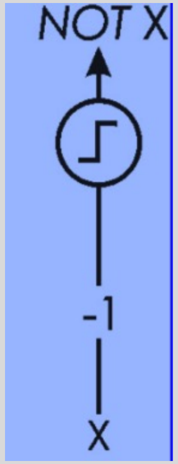
Introduction of AI (CS103)- 06
Overview of Machine Learning and Supervised Learning

Jimmy Liu 刘江

2023-10-27

Correction： M-P Model –”Not”





x_1	θ	w_1	$f(\cdot)$
0	0	-1	$G(z)$
1	0	-1	$G(z)$

$$G(z) = \begin{cases} 1 & if\ z \geq 0 \\ 0 & otherwise \end{cases}$$

$z = w_1 x_1 - \theta$	Output $y = \mathrm{G}(z)$	NOT A
0 * (-1) - 0 = 0	1	1
1 *(-1) – 0 = -1	0	0

2022 Mid-Term Test
– Lecture 7 Will Be Our Test Lecture

南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Course Name: **Introduction of Artificial Intelligence**

Dept.: **Department of Computer Science and Engineering**

Exam Duration: **2 hours** **Exam Paper Setter:** **Liu Jiang**

Question No.	1	2	3	4	5	6	7	8	9	10
Score	2	4	5	4	5					

This exam paper contains ___5___questions and the score is ___20__ in total. (Please hand in your exam paper, answer sheet, and your scrap paper to the proctor when the exam ends.)

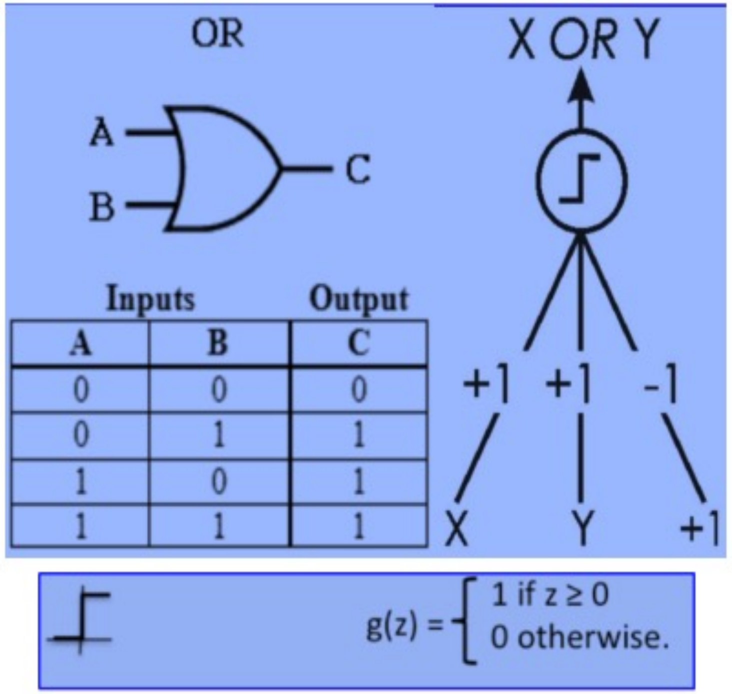
1. Please list the 4 abilities of the concept "Intelligence". **(2 marks)**

2. What is reinforcement learning? Please draw the typical framing of a reinforcement learning scenario and list 5 key components. **(4 marks)**

3. Please draw the artificial neuron of a Perceptron and write down the Perceptron learning algorithm step by step (the **Pseudo code** or **description**). **(5 marks)**

4. Please list 2 key supporting algorithms behind ALPHAGO and how the learning algorithm supports the success of ALPHAGO in the Go Game against human player. **(4 marks)**

5. Prove the following MCP neuron is a "OR" Artificial Neuron. **(5 marks)**



Content of Lecture 6

1. Review of Lecture 5
2. Overview of Machine Learning
3. Definition and Overview of Supervised Learning
4. Linear Regression
5. Logistic Regression
6. Perceptron

Overview of Machine Learning

What is Machine Learning?

- Machine learning (ML) is a concept proposed by Arthur Samuel, known as the "father of machine learning," in 1956. He believed that machine learning is a research field that endows computers with the ability to learn without directly programming them to solve problems.

- From a disciplinary perspective, machine learning is an interdisciplinary field that encompasses probability theory, statistics, calculus, computer science, and other fields.

- Machine learning is dedicated to studying how to enable computers to simulate or implement human learning methods to acquire new knowledge or skills while reorganizing existing knowledge structures to continuously improve their learning efficiency and performance.

Overview of Machine Learning

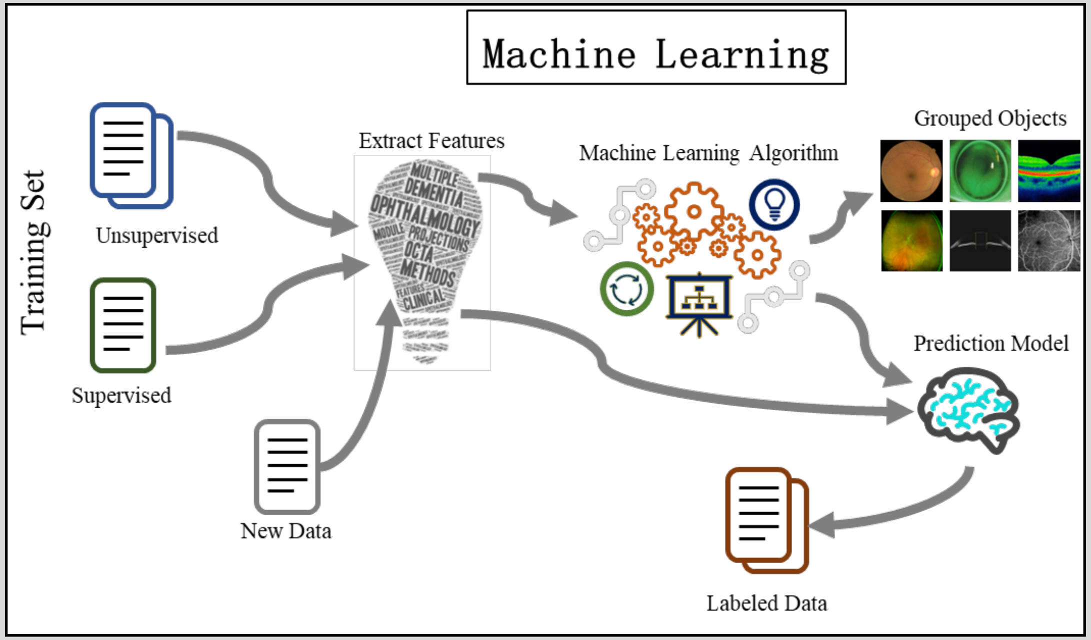


Figure 1 The Learning process of machine learning

Learning Process of Machine Learning

The learning process of machine learning is mainly composed of two parts: training and testing (validation, etc. can also be included in testing).

Training: The computer uses a large amount of data (training set) through supervised or unsupervised means to train and model, and learns the potential distribution rules in the data.

Testing: The model constructed on the training data samples is used to classify or predict the corresponding output (labeled data) on new data samples (test set) that have not been seen during training

Overview of Machine Learning

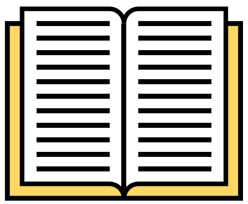
Dataset of Machine Leaning

In machine learning literature, you will often come across the terms training set, validation set, and test set. These three terms are defined based on their specific roles in machine learning:

Training set: A set of data samples used to build a model. It can be compared to the content of the CS 103 course textbook and the knowledge taught by a teacher in class (Learning knowledge based on textbooks and teaching content).

Validation set: A set of data samples used to adjust the hyperparameters of a model or to make an initial evaluation of the model's generalization ability. It can be compared to course CS 103 assignments or class test (Assignments can provide feedback on knowledge mastery). It is an optional dataset, and sometimes there may not be a validation set.

Test set: A set of data samples used to test the generalization ability of a model. It can be compared to CS 103 course exams (Testing students' course abilities based on new question types derived from learning content).



Training Set

Validation Set

Testing Set

Overview of Machine Learning

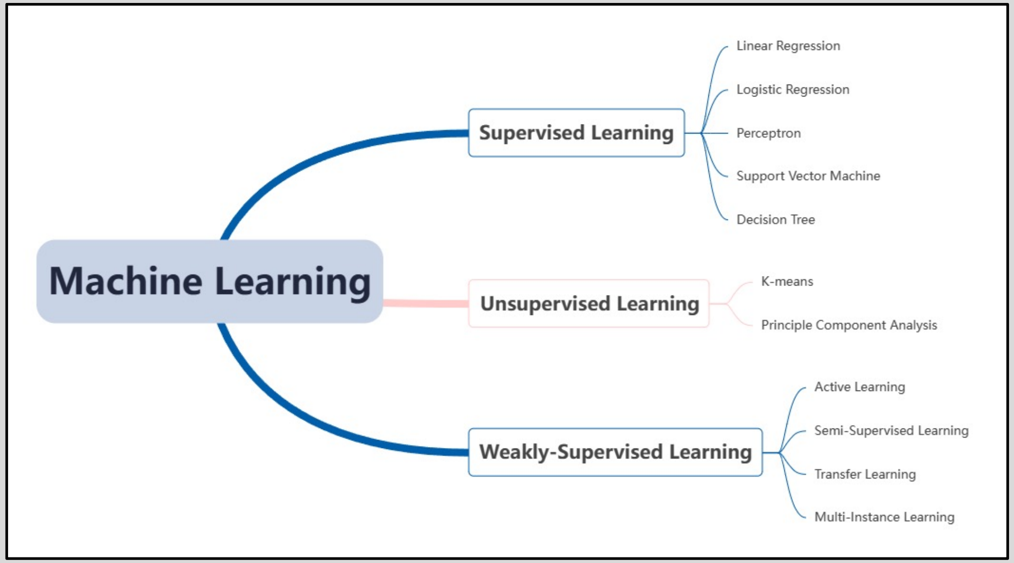


Figure 2 Three types of Machine Learning

Algorithms of Machine Learning

There are multiple classification methods for machine learning algorithms. In this course, machine learning algorithms are classified into three types based on the information provided by the training samples and the feedback method: **Supervised Learning**, **Unsupervised Learning**, and **Weakly-Supervised learning**. In this section and future sections, some classic algorithms of these types will be introduced

Overview of Machine Learning

AI vs ML vs DL

Artificial Intelligence is a goal that various algorithms are currently striving to achieve, which is to create an intelligent agent with intelligence. Machine learning is one of the means to achieve artificial intelligence. The currently hotly developed deep learning algorithm is a subfield in the field of machine learning, mainly implemented using neural networks. The previous video explains the relationship between the three of them (Correction of the video: deep learning has become the mainstream application algorithm of artificial intelligence).

Q1:Selecting the Machine Learning Application from the following choices.

A. Garbage spam recognition ；
B. Prediction of the House Price of One Area ；
C. Given some numbers, calculate the summary of these numbers. ；
D. Given a final value and some conditions, let the computer deduce the initial value;
E. Given some pictures, let the computer recognize the category of the object in the picture ；

Content of Lecture 6

Definition and Overview of Supervised Learning

Definition of Supervised Learning

- Supervised Learning is so called because during the training process, there are label samples to prompt the correctness of the model output. The model adjusts its learning parameters according to the correctness informed by the label data, and reaches the final performance after continuous adjustment. This process is supervised learning or supervised training.
- In mathematical definition, given that the data sample feature x has a corresponding label y to guide the mathematical modeling to obtain the function model $f(x)$ or $p(y|x)$, this type of learning algorithm is called supervised learning algorithm.
- Supervised learning can be roughly divided into two categories: classification algorithms and regression algorithms, depending on whether the labels in the data are discrete or continuous.

Definition and Overview of Supervised Learning

Instance of Supervised Learning

If the features of the slit lamp or OCT image are regarded as input x, and the severity level of the patient' s cataract is regarded as label y, then during the training process, the algorithm outputs a category $\hat{y}$, By comparing with label y, it can be known whether the algorithm has outputted the correct result. Based on this result, the algorithm can decide on parameter updates. You can specify how many steps to train or use some evaluation indicators (such as how much the classification accuracy rate has reached) to end the training and get the final model.

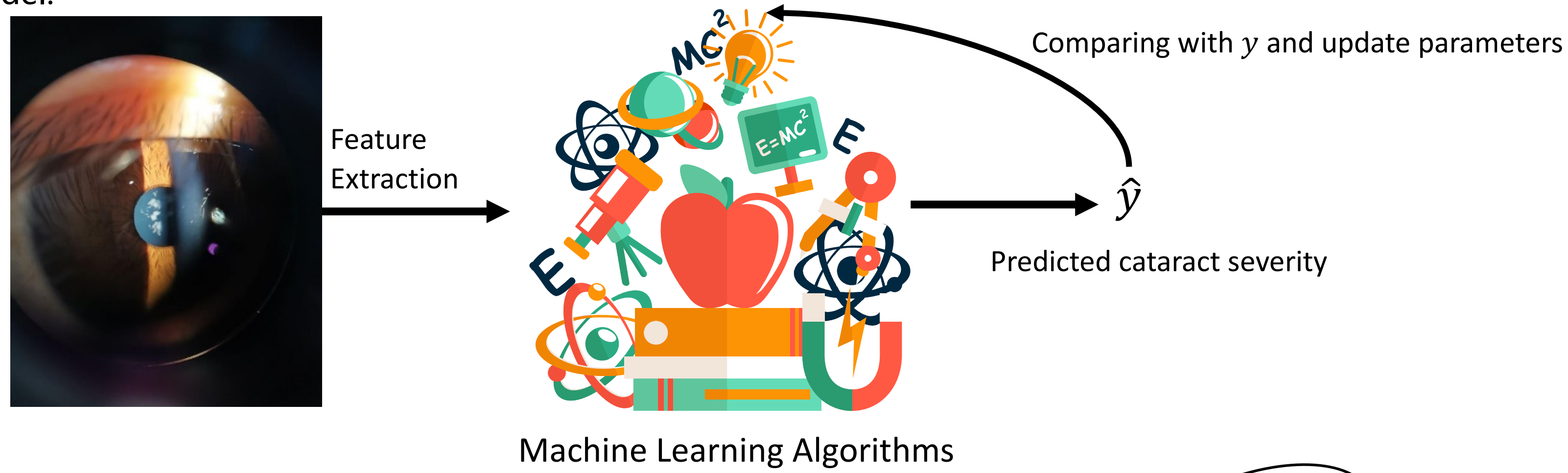
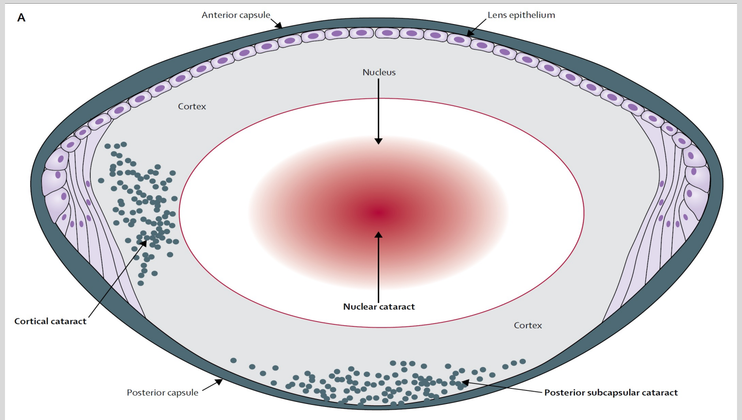
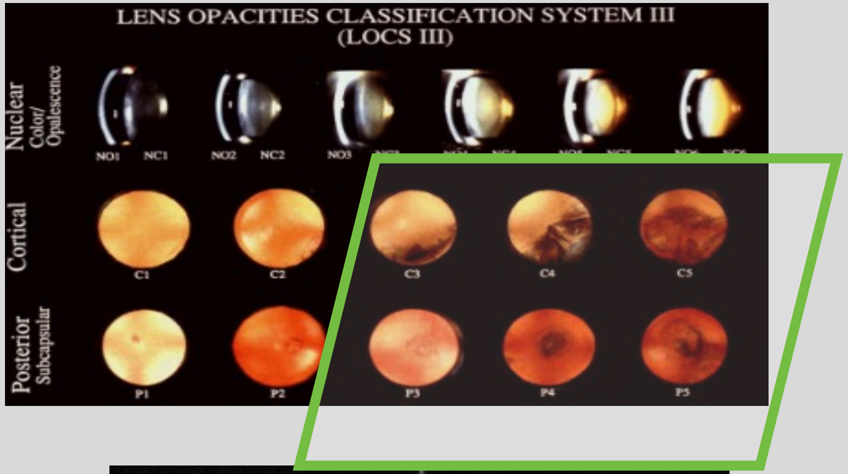
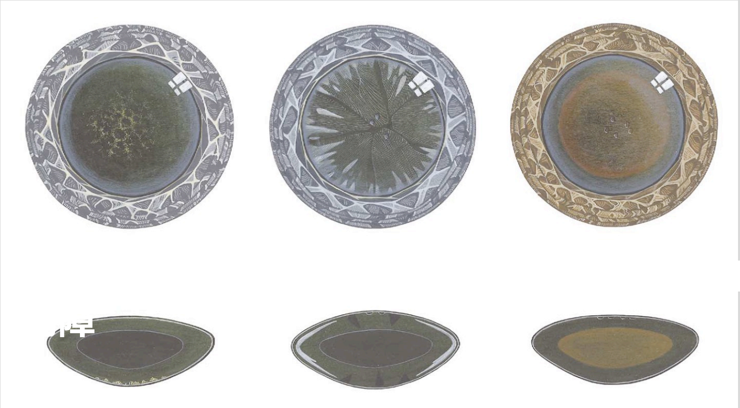


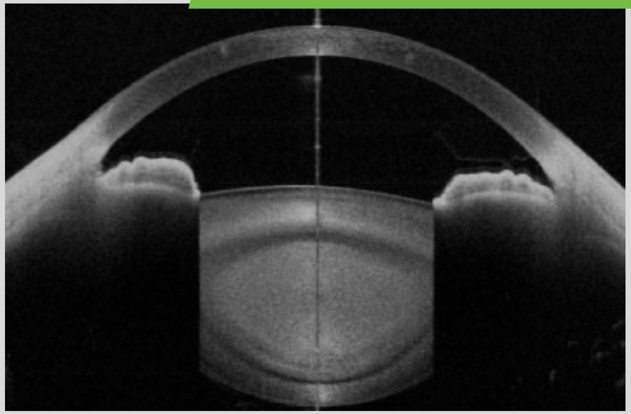
Figure 3 The procedure for the prediction of the severity of cataract

Example：Cataract Classification









Definition of Supervised Learning

Algorithm of Supervised Learning

Supervised learning algorithms have a wide range of applications in the field of artificial intelligence, and many applications are built based on supervised learning algorithms. This course will detail several classic supervised learning algorithms as shown in Figure 5, including linear regression, logistic regression, perceptron, support vector machine, decision tree, and ensemble learning.”

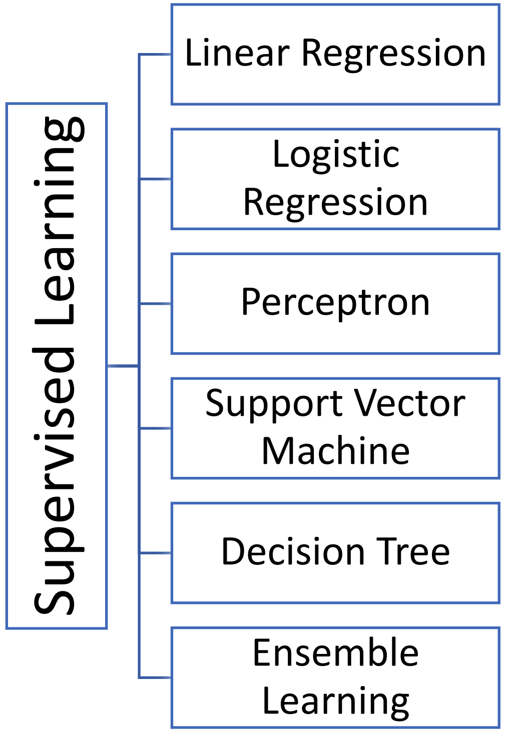
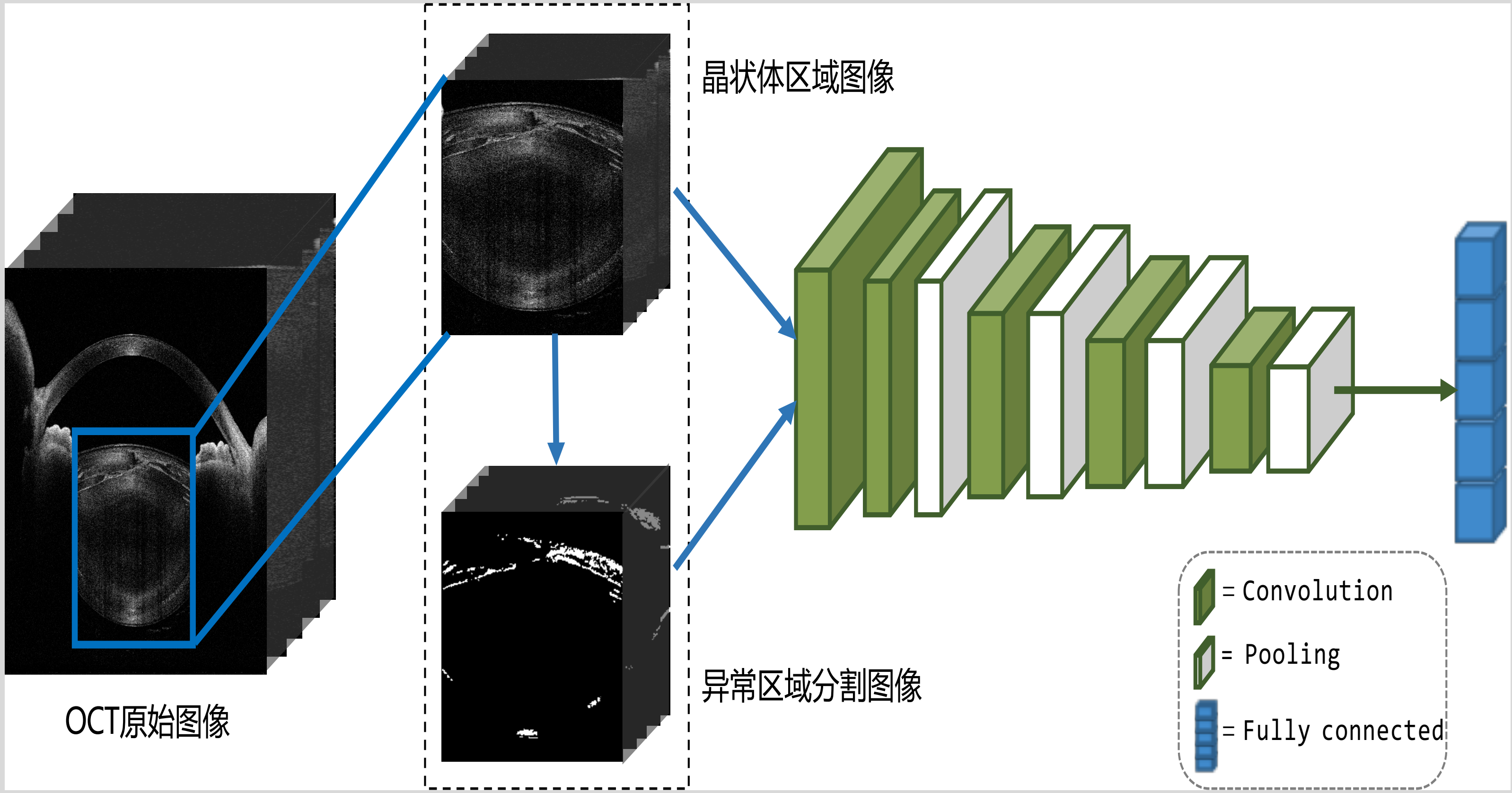


Figure 5 Classic supervised learning algorithms

Cataract Classification through Supervised Learning

（TOMEY CASIA-II）



Process of Supervised Learning

According to the case and the definition of supervised learning, supervised learning can be summarized as the flowchart shown in the following figure. First, the input data is converted into features needed by the algorithm through some methods. This process is represented by interpretation in the figure to represent the feature extraction process. This process will be hidden in the algorithm in the later course of deep learning end-to-end training. After extracting features, the algorithm is trained on the training set to obtain a model with strong performance. Finally, there may be some post-processing operations on the model output to produce better results.

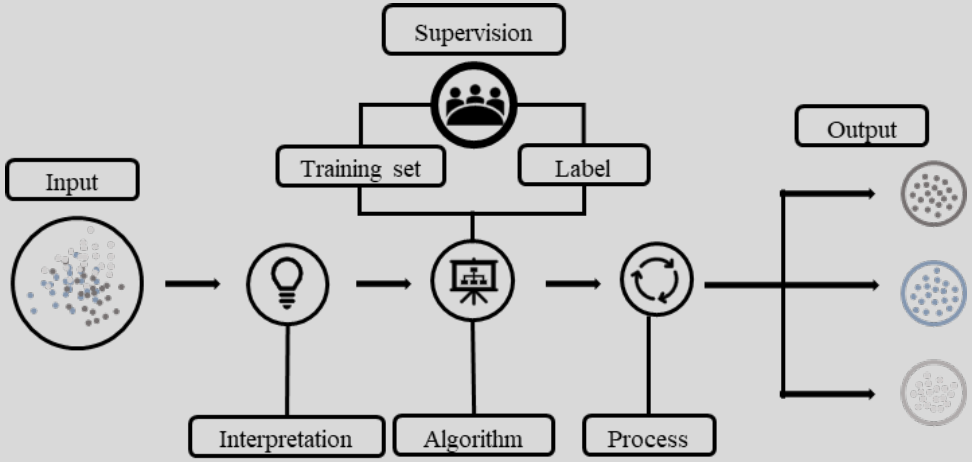
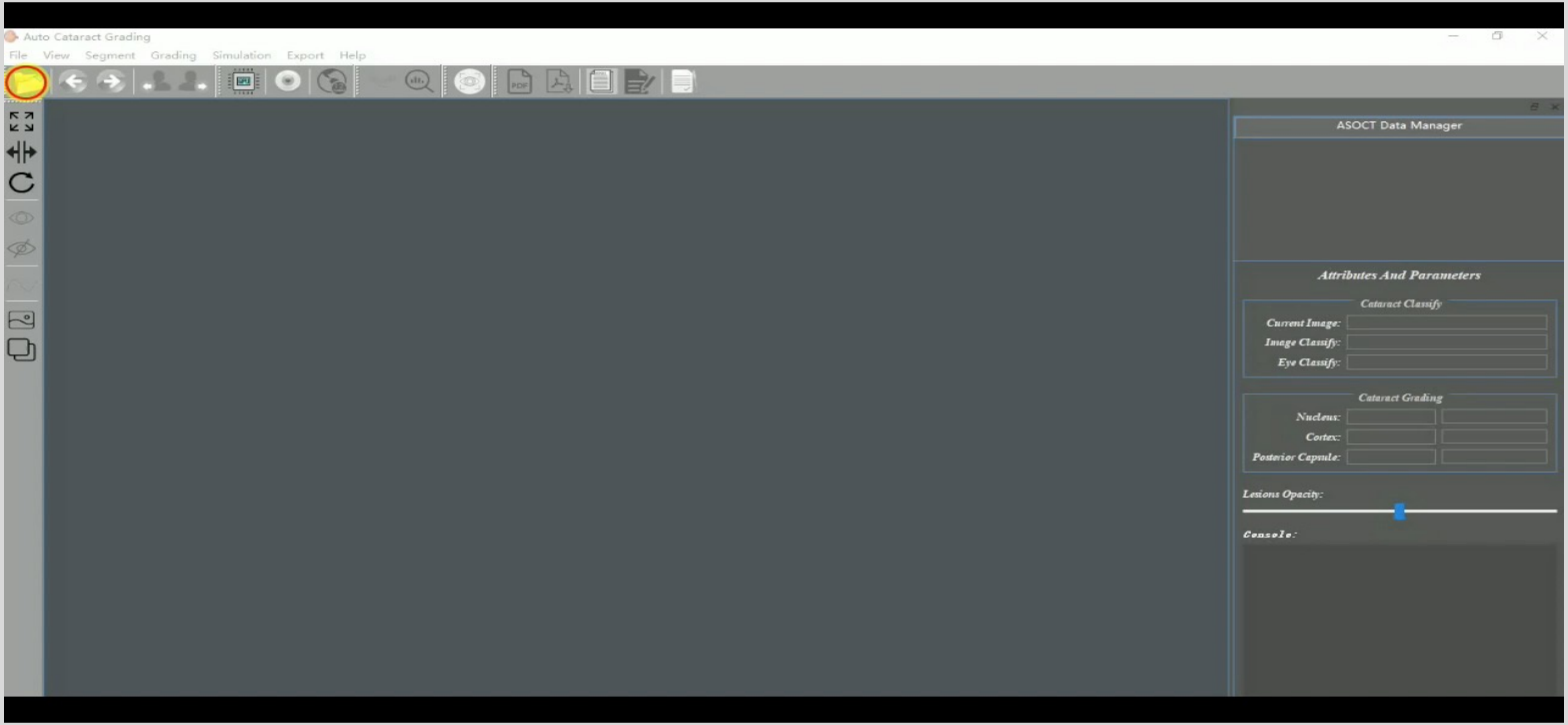


Figure 4 The basic process of the supervised learning.

Cataract Classification Process



Q2: Please think about the drawbacks of Supervised Learning

Content of Lecture 6

Classic Supervised Learning Algorithm-Linear Regression

Definition of Linear

- Linear is that given one mapping $f(x) = C$, where $f(x)$ is a linear mapping, then it is linear, otherwise nonlinear.
- Linear mapping must satisfy additivity and homogeneity :

  Additivity: $f(x + y) = f(x) + f(y)$
- Homogeneity: $f(ax) = af(x)$
- The function $f(x) = kx$ is one linear mapping, and figure 6 is one sample of it.

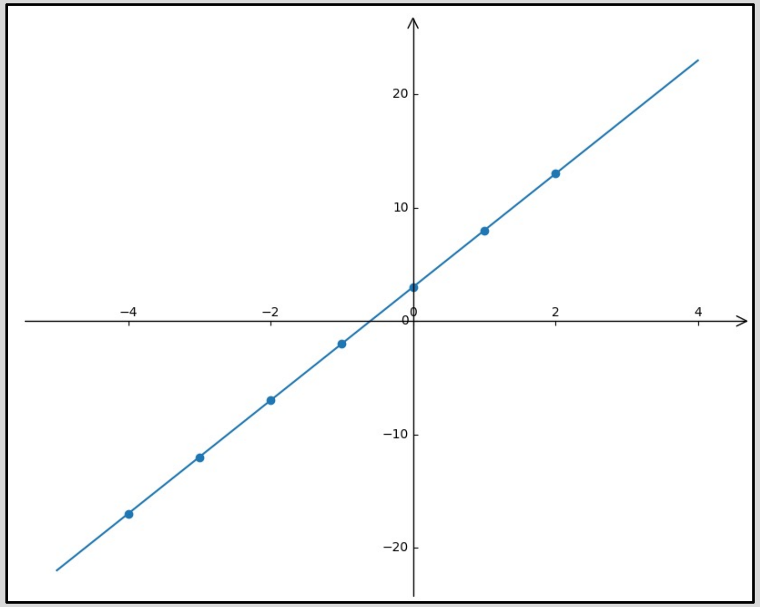


Figure 6 Linear mapping

Classic Supervised Learning Algorithm-Linear Regression

Definition of Linear Regression

The Linear Regression model is a type of supervised learning algorithm mainly used for regression problems. It obtains the final prediction results by linearly combining the input samples, so it can directly observe the direct impact on the output in the samples. Its mathematical definition is as follows:

Given one sample $\boldsymbol{x} = [x_1, x_2, x_3, \dots, x_d]$ with dimension d, we can use one vector $\boldsymbol{w} \in \boldsymbol{R}^d$ and one bias b to represent the target function:

$$f(x) = w_1x_1 + w_2x_2 + \cdots + w_dx_d + b = \boldsymbol{w}^T\boldsymbol{x} + \mathrm{b}$$

The computation graph of the linear regression is shown in Fig.7.

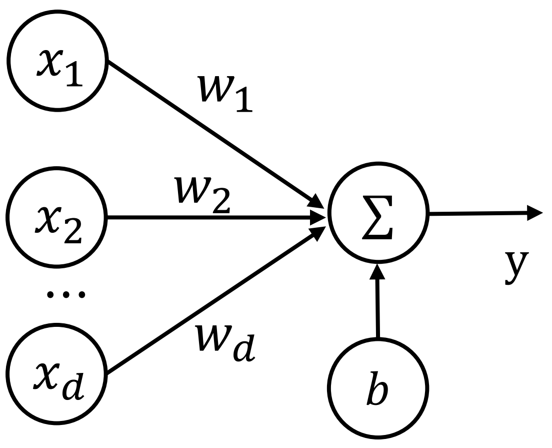


Figure 7 Computation graph of Linear Regression

Classic Supervised Learning Algorithm-Linear Regression

Med

Linear Regression application example

To facilitate the understanding of the linear regression algorithm, myopia degree prediction is used as an example to demonstrate the linear regression algorithm here. The degree of myopia is related to many clinical features, among which naked eye visual acuity, equivalent spherical degree, and eye axis length are three important features. The table below gives samples of 5 myopic patients, each sample contains the above three features, so it can be represented as $x_i = [x_{i1}, x_{i2}, x_{i3}]$. According to the definition of linear regression, the degree of myopia of each sample can be represented by the equation $y = f(x_i) = w_1 x_{i1} + w_2 x_{i2} + w_3 x_{i3} + b = \boldsymbol{w}^T \boldsymbol{x} + b$. Here, the first three samples (1-3) are selected as training set samples, and the last two (4-5) are set as validation set samples.

Table 1 Samples of degree of myopia

Index of Sample i	Naked eye visual acuity x_{i1}	*Equivalent spherical degree* x_{i2}	Eye axis length x_{i3}	Degree of myopia $y = f(x_i)$
1	4.5	-3	25	12.5
2	4	-4	25	12.6
3	4.5	-2.5	25	12.4
4	4.5	-2	20	10.3
5	3.5	-10.5	30	15.8

Classic Supervised Learning Algorithm-Linear Regression

Linear Regression application example

After achieving the equation $y=\boldsymbol{w}^T \boldsymbol{x}+b$, we do not know the value of $\boldsymbol{w}$ and b, how to learn these two parameters is the target of linear regression. In machine learning, the loss function or objective function is often used to guide the learning of parameters. The loss function is also a function used to measure the error between the predicted value and the true value. In linear regression, mean square error is commonly used as the loss function. The process of learning from samples is to optimize (update) parameters by some methods so that the output obtained by the model based on samples has the smallest loss function value which is also the smallest mean square error.

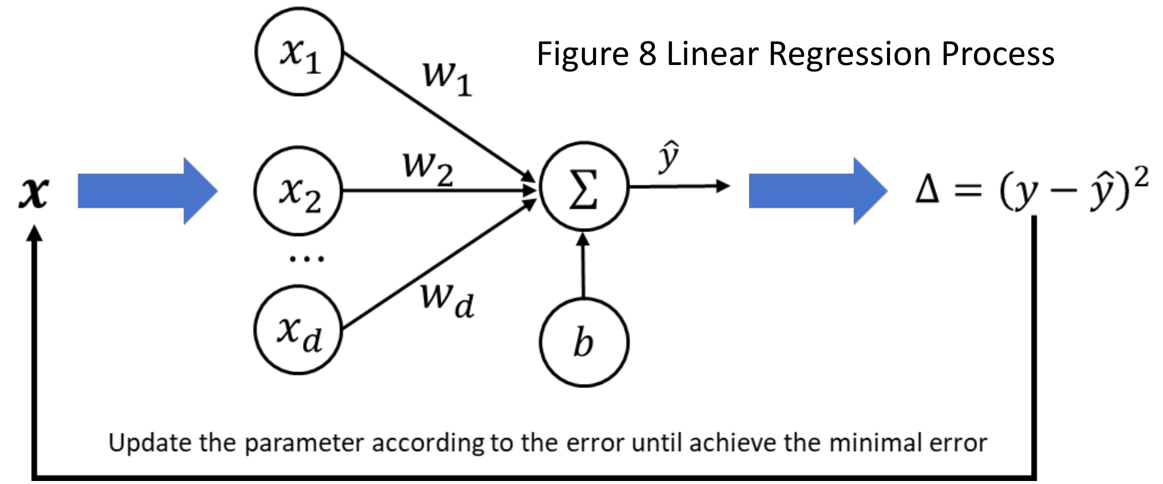


Figure 8 Linear Regression Process

Classic Supervised Learning Algorithm-Linear Regression

Linear Regression application example

- How to update the parameters? The commonly used methods are the least squares method and the gradient descent method. The gradient descent method will be detailed when introducing deep learning later. Here, only the method of gradient descent is used to explain how to update parameters. The gradient descent method mainly uses the formula $w' = w - \eta\partial L/\partial w$to update parameters, where η is the learning rate, and $\partial L/\partial w$ is the partial derivative of the parameter, also known as the gradient. The learning rate is one hyperparameter set by manual.
- We can initialize the $\boldsymbol{w}$ and b as $\boldsymbol{w} = [0.3, -0.1, 0.5], b = 1$, and set the learning rate η to $\eta = 0.0001$. Then according to the mean square error function, the sum of the error in the training set is 20.5625, and the summary of the mean square error in the validation set is 10.3525.

Classic Supervised Learning Algorithm-Linear Regression

The application example of Linear Regression

- According to the mean square error formula, the gradients of parameters $\boldsymbol{w}$ and $b$can be obtained by taking partial derivatives, which are [68.15,-49.4,392.5] and 19.7 respectively. According to the formula of gradient descent method, the updated parameters $\boldsymbol{w}' = [0.293185, -0.09506, 0.46075]$ and $b' = 0.99803$ are achieved.
- By substituting the updated parameters into the linear regression model and recalculating the sum of errors, it can be found that the sum of errors on the training set has dropped to 7.5903, and on the validation set it has dropped to 3.1151. It can be seen that this method can effectively learn the parameters needed by the model.

Q3: Could linear regression be used in classification tasks?

Content of Lecture 6

Classic Supervised Learning- Logistic Regression

Binary classification Problem

- From the name, binary classification is to divide the input into two categories. Its definition is that given the input feature x, the corresponding classification output $y \in \{0,1\}$ can be obtained, where 0 is the negative class (negative sample), and 1 is the positive class (positive sample).
- Taking the discrimination between illness and health as an example, we can assume that health is a negative class 0, and illness is a positive class 1. Then, judging the input sample data as healthy (0) or sick (1) through some algorithm is a process of binary classification.
- The logistic regression that I am going to talk about now is a linear model used to solve classification problems.

Q4:How to convert the continuous value to discreate.

Classic Supervised Learning- Logistic Regression

Logistic Regression

- Logistic Regression is a machine learning algorithm to solve the binary (0 or 1) classification problem by estimating the probability of an event.

- When applying the linear regression model to solve the classification model, a nonlinear decision function needs to be introduced to construct the mapping relationship between the continuous real number output of the model and the classification label.

Classic Supervised Learning- Logistic Regression

Logistic Regression

- In binary classification problems, a nonlinear function $g(\cdot)$ is usually used as a decision function (also known as an activation function) to predict the posterior probability $p(y=1|x)$.
- If $f(\boldsymbol{x}, \boldsymbol{w}) = \boldsymbol{w}^T\boldsymbol{x} + b$ represents the linear regression model, then $p(y=1|x) = \mathrm{g}(\mathrm{f}(\boldsymbol{x}, \boldsymbol{w}))$ can be used to represent the logistic regression model. Figure 9 is the calculation graph of the logistic regression model.

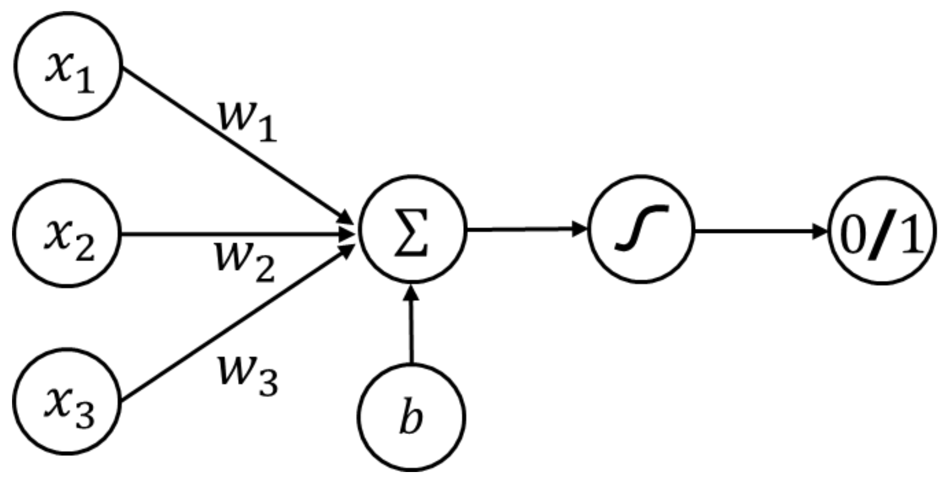


Figure 9 Computation graph of Logistic Regression

Classic Supervised Learning- Logistic Regression

Logistic Regression

- In logistic regression, the Sigmoid function is commonly used to map the continuous value to the interval (0,1). The formulation of the Sigmoid function is as follows:

$$y = \frac{1}{1 + e^{-x}}$$

- According to the formulation of sigmoid and the formulation of logistic regression, we can rewrite the logistic model as $\boldsymbol{w}^T\boldsymbol{x} + b = ln\frac{p(y=1|x)}{p(y=0|x)}$, where $ln\frac{p(y=1|x)}{p(y=0|x)}$ is log-odds or logit. Thus we call the algorithm as Logistic regression.

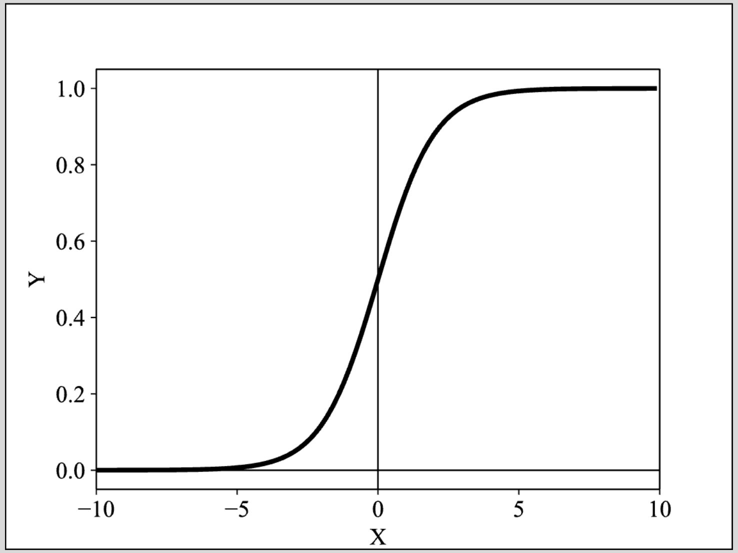


Figure 10 Sigmoid function

Classic Supervised Learning- Logistic Regression

Med

Logistic regression application example

To facilitate the understanding of the logistic regression algorithm, myopia recognition is used as an example to demonstrate the logistic regression algorithm here. The degree of myopia is related to many clinical features, among which naked eye visual acuity, equivalent spherical degree, and eye axis length are three important features. The table below gives samples of 5 myopic patients, each sample contains the above three features, so it can be represented as $x_i = [x_{i1}, x_{i2}, x_{i3}]$, $y \in \{0,1\}$ is the label of sample , 0 represent not myopia , 1 represent the patient has myopia。The first three samples (1-3) are selected as training set samples, and the last two (4-5) are set as validation set samples

Table 2 Samples of myopia

Index of Sample i	Naked eye visual acuity x_{i1}	*Equivalent spherical degree* x_{i2}	Eye axis length x_{i3}	Is myopia? $y = f(x_i)$
1	5.1	-1.25	23.06	0
2	4.6	-1.13	23.24	1
3	4.6	-1.13	24.62	1
4	5	-1.13	23.06	0
5	4.9	-0.88	23.67	1

Classic Supervised Learning- Logistic Regression

Logistic regression application example

- Different from linear regression, the cross entropy loss function :L = -∑ yi * log(pi) is usually as the objective function here.
- The parameters $\boldsymbol{w}$ and b are initialized to $\boldsymbol{w}$=[1,1,1], b=1. Here we set the learning rate η as η=0.1. Now, the sum of the loss on the training set is 9.3034, and the sum of the loss on the validation set is 13.9650.

Classic Supervised Learning- Logistic Regression

Logistic regression application example

- According to the cross-entropy loss formula, the gradients of parameters $\boldsymbol{w}$ and b can be obtained by taking partial derivatives, which are [1.7000,-0.4167,7.6867] and 0.3333 respectively. According to the formula of gradient descent method, the updated parameters $\boldsymbol{w}' = [0.8300, 1.04167, 0.23133]$ and bias $b' = 0.96667$ are obtained.
- By substituting the updated parameters into the logistic regression model and recalculating the sum of errors, it can be found that the sum of losses on the training set has dropped to 3.0775, and on the validation set it has dropped to 4.6471.

Classic Supervised Learning Algorithm-Other Regression Algorithms

Other Regression Algorithms

- In addition to the linear regression and logistic regression algorithms introduced just now, there are some classic regression algorithms, such as polynomial regression, stepwise regression, ridge regression, lasso regression, elastic network regression, etc. The following figure summarizes the characteristics of these regression algorithms.

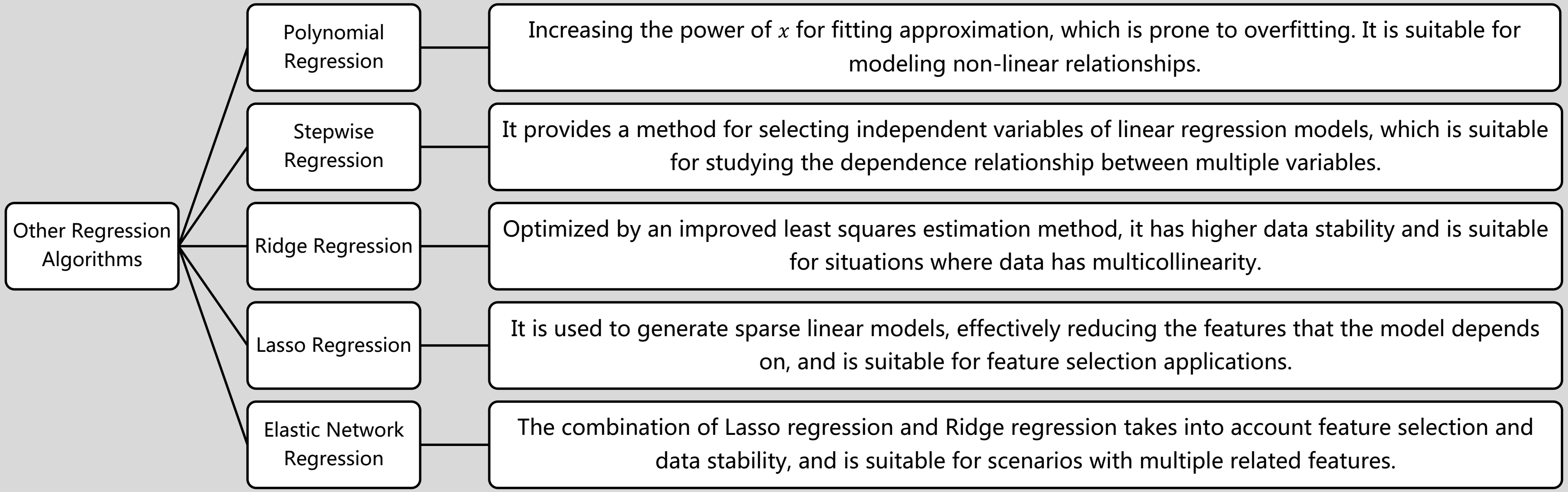


Figure 11:Other Regression Algorithms

Content of Lecture 6

1 Review of Lecture 5

2 Overview of Machine Learning

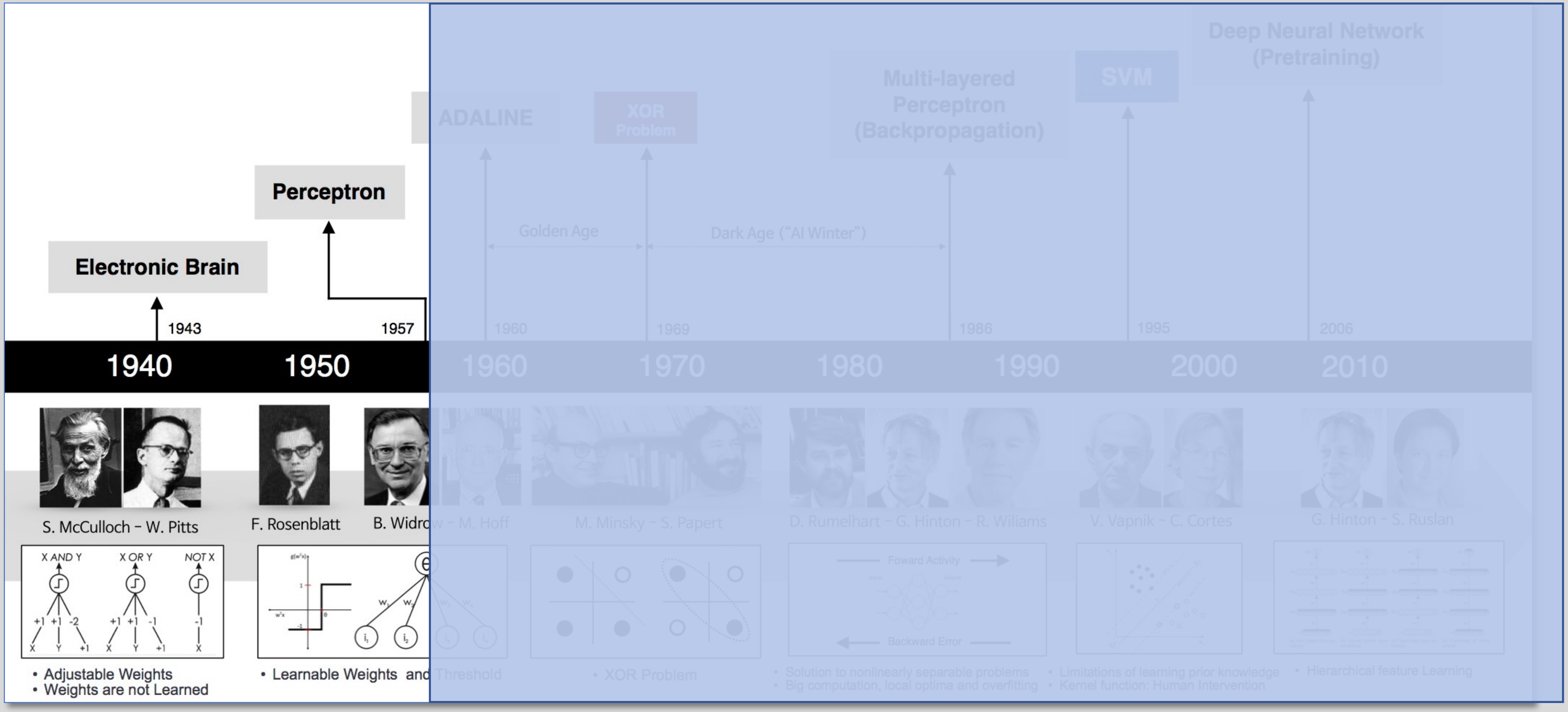
3 Definition and Overview of Supervised Learning

4 Linear Regression

5 Logistic Regression

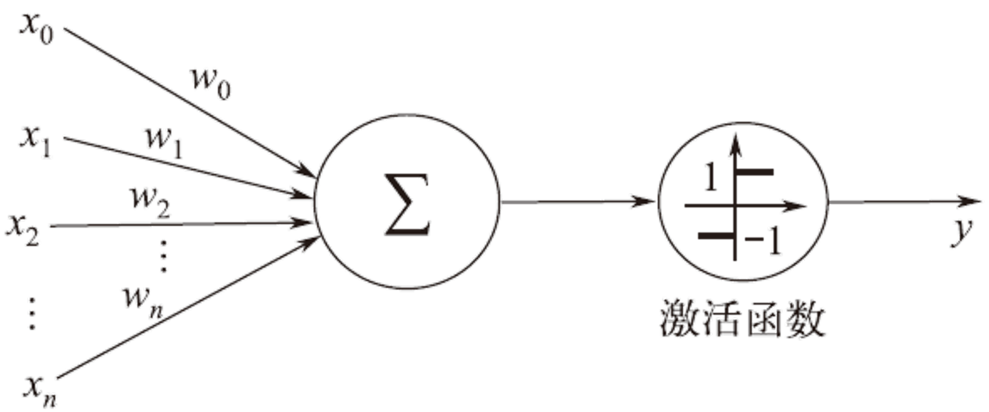
6 Perceptron

AI Development Milestone- Perceptron



Perceptron

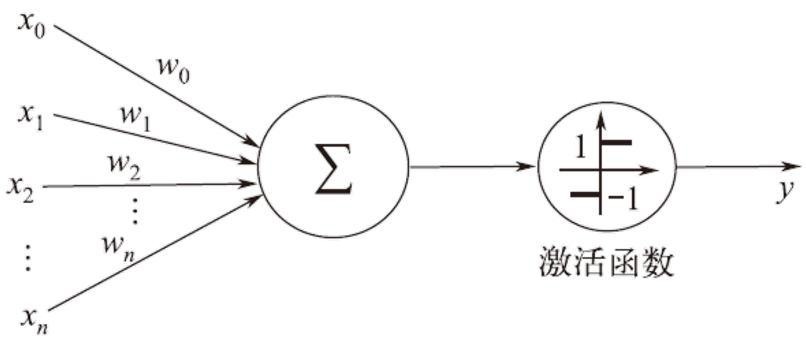
Perceptron was invented by Frank Rosenblatt in 1957, as a classic neural network structure and widely used linear classifier.



PERCEPTION

They were one-to-one corresponding to Bio-Neuron:

Input --- Dendrite
Weight --- Synapse
Activation Function --- Nucleus



PERCEPTION

Perception can be seen as simply as Binary Classification Model

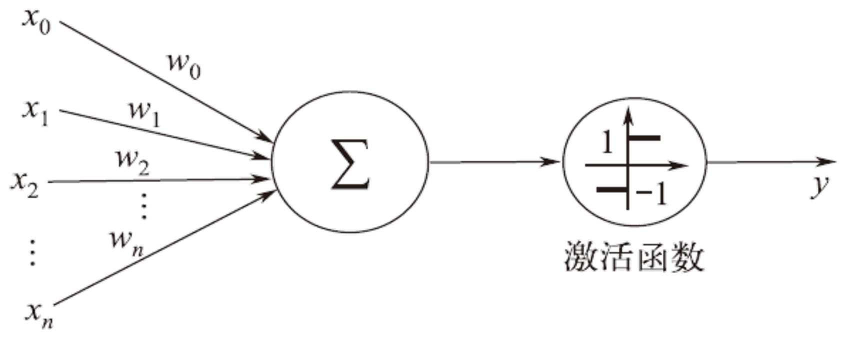
Activation Function

$$\hat{y} = sgn(w^T x)$$

Objective value Learnable Weight Input Feature

PERCEPTION

Perception can be seen as simply as Binary Classification Model



$$w \longleftarrow w + yx$$

PERCEPTION

Algorithm：Perception Training Algorithm

Input：Dataset $D\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ Maximum Training Round E

1: Initialize：w←0 , e← 0, k ←0

2: For e to E do

3: Randomly mess the Dataset D

4: i ←0

5: while (i < n) do

6: $\hat{y}_i \longleftarrow w_k^T x_i$

7: if $\hat{y}_i y_i <= 0$ then

8: $\omega_{k+1} \longleftarrow \omega_k + y_i x_i$

Output: k w

Myopia Example

We plan to design a diagnose algorithm to judge whether the patients have severe myopia based on three features shown here.

Table 3-4 Myopia sample dataset

No	Naked eye visual acuity x_{i1}	*Equivalent spherical degree* x_{i2}	Eye axis length x_{i3}	Severe myopia？ (y)
1	3.7	–10.5	27.49	+1
2	4.8	0.13	21.75	–1
3	4.9	1.25	22.31	–1
4	3.8	–6.38	24.71	+1
5	5.0	–0.38	22.79	–1

Myopia Example

Q1 ：Take the first three as training sample and following two as testing sample, set the maximum training round to 4.

How does the training process work?

Table 3-4 Myopia sample dataset

No	Naked eye visual acuity x_{i1}	*Equivalent spherical degree* x_{i2}	Eye axis length x_{i3}	Severe myopia？（y）
1	3.7	–10.5	27.49	+1
2	4.8	0.13	21.75	–1
3	4.9	1.25	22.31	–1
4	3.8	–6.38	24.71	+1
5	5.0	–0.38	22.79	–1

Myopia Example

Algorithm：Perception Training Algorithm
Input：Dataset $D\{(x_1,y_1),(x_2,y_2),...,(x_n,y_n)\}$ Maximum Training Round E
1: Initialize：w←0 , e← 0, k ←0
2: For e to E do
3: Randomly mess the Dataset D
4: i ←0
5: while (i < n) do
6: $\hat{y}_i \longleftarrow w_k^T x_i$
7: if then
8: $\omega_{k+1} \longleftarrow \omega_k + y_i x_i$

Output: k w

No	Naked eye visual acuity x_{i1}	*Equivalent spherical degree* x_{i2}	Eye axis length x_{i3}	Severe myopia？（y）
1	3.7	–10.5	27.49	+1
2	4.8	0.13	21.75	–1
3	4.9	1.25	22.31	–1
4	3.8	–6.38	24.71	+1
5	5.0	–0.38	22.79	–1

Home Work 6: Please Verify the Table

Iter	$\boldsymbol{x}$	$\boldsymbol{y}$	$\hat{y}$	$\boldsymbol{w}$
0	(3.7, –10.5, 27.49)	+1	0	(0, 0, 0)
1	(4.8, 0.13, 21.75)	–1	+1	(3.7, –10.5, 27.49)
2	(4.9, 1.25, 22.31)	–1	+1	(–1.1, –10.63, 5.74)
3	(4.8, 0.13, 21.75)	–1	–1	(–6., –11.88, –16.57)
4	(3.7, –10.5, 27.49)	+1	–1	(–6., –11.88, –16.57)
5	(4.9, 1.25, 22.31)	–1	+1	(–2.3, –22.38, 10.92)
6	(3.7, –10.5, 27.49)	+1	–1	(–7.2, –23.63, –11.39)
7	(4.9, 1.25, 22.31)	–1	+1	(–3.5, –34.13, 16.1)
8	(4.8, 0.13, 21.75)	–1	–1	(–8.4, –35.38, –6.21)
9	(3.7, –10.5, 27.49)	+1	+1	(–8.4, –35.38, –6.21)
10	(4.9, 1.25, 22.31)	–1	+1	(–8.4, –35.38, –6.21)
11	(4.8, 0.13, 21.75)	–1	–1	(–8.4, –35.38, –6.21)

When do weights stop updating? How does this set of weights work in the test sample?

```
Algorithm：Perception Training Algorithm
Input：Dataset D{(x_1,y_1),(x_2,y_2),...,(x_n,y_n)} Maximum Training Round E
Initialize：w←0 , e← 0, k ←0
For e to E do
        Randomly mess the Dataset D
        i ←0
        while (i < n) do
            ŷ_i ⟵ w_k^T x_i
            if ŷ_i y_i <= 0 then
                ω_{k+1} ⟵ ω_k + y_i x_i

Output: k w
```

When do weights stop updating? How does this set of weights work in the test sample?

Iter	x	y	$\hat{y}$	w
0	(3.7, –10.5, 27.49)	+1	0	(0, 0, 0)
1	(4.8, 0.13, 21.75)	–1	+1	(3.7, –10.5, 27.49)
2	(4.9, 1.25, 22.31)	–1	+1	(–1.1, –10.63, 5.74)
3	(4.8, 0.13, 21.75)	–1	–1	(–6., –11.88, –16.57)
4	(3.7, –10.5, 27.49)	+1	–1	(–6., –11.88, –16.57)
5	(4.9, 1.25, 22.31)	–1	+1	(–2.3, –22.38, 10.92)
6	(3.7, –10.5, 27.49)	+1	–1	(–7.2, –23.63, –11.39)
7	(4.9, 1.25, 22.31)	–1	+1	(–3.5, –34.13, 16.1)
8	(4.8, 0.13, 21.75)	–1	–1	(–8.4, –35.38, –6.21)
9	(3.7, –10.5, 27.49)	+1	+1	(–8.4, –35.38, –6.21)
10	(4.9, 1.25, 22.31)	–1	+1	(–8.4, –35.38, –6.21)
11	(4.8, 0.13, 21.75)	–1	–1	(–8.4, –35.38, –6.21)

We can see from Table that the weight for the model update when the output is incoherent with the label.

If the weight can accurately predict the data in the training set, it stopped updating.(In step 7, it updates. After step 8, it stops updating)

In that way we can get the new final weight vector as:

$$\omega = (-8.4, -35.38, -6.21)$$

Myopia Example

We can get the new final weight vector $\omega = (-8.4, -35.38, -6.21)$

Placing the weight in the testing part, we can see this vector is the optimal one. It can be found that this set of parameters can distinguish the samples in the test set, so this set of parameters is a better parameter to predict whether the patient has severe myopia.

$\boldsymbol{x}$	$\hat{\boldsymbol{y}}$	$\boldsymbol{y}$
(3.8, –6.38, 24.71)	+1	+1
(5.0, –0.38, 22.79)	–1	–1

Home Work 6: Please Verify the Table

Iter	$\boldsymbol{x}$	y	$\hat{y}$	$\boldsymbol{w}$
0	(3.7, –10.5, 27.49)	+1	0	(0, 0, 0)
1	(4.8, 0.13, 21.75)	–1	+1	(3.7, –10.5, 27.49)
2	(4.9, 1.25, 22.31)	–1	+1	(–1.1, –10.63, 5.74)
3	(4.8, 0.13, 21.75)	–1	–1	(–6., –11.88, –16.57)
4	(3.7, –10.5, 27.49)	+1	–1	(–6., –11.88, –16.57)
5	(4.9, 1.25, 22.31)	–1	+1	(–2.3, –22.38, 10.92)
6	(3.7, –10.5, 27.49)	+1	–1	(–7.2, –23.63, –11.39)
7	(4.9, 1.25, 22.31)	–1	+1	(–3.5, –34.13, 16.1)
8	(4.8, 0.13, 21.75)	–1	–1	(–8.4, –35.38, –6.21)
9	(3.7, –10.5, 27.49)	+1	+1	(–8.4, –35.38, –6.21)
10	(4.9, 1.25, 22.31)	–1	+1	(–8.4, –35.38, –6.21)
11	(4.8, 0.13, 21.75)	–1	–1	(–8.4, –35.38, –6.21)

Introduction of AI (CS103)- 06 Overview of Machine Learning and Supervised Learning

Jimmy Liu 刘江

2023-10-27